J. Mol. Biol. (1963) **6**, 1–15

Relationships between the Ribonucleic Acid and Protein of Some Plant Viruses

R. H. Symons,† M. W. Rees, Margaret N. Short and R. Markham

Agricultural Research Council Virus Research Unit, Huntingdon Road, Cambridge, England

(*Received 18 August 1962*)

Six strains of turnip yellow mosaic virus‡ and one each of wild cucumber mosaic virus and turnip crinkle virus have been analysed for the amino acid composition of their proteins, the base ratios and the products of pancreatic ribonuclease digestion of their RNA.

The results for TYMV show that the six strains fall into two well-defined groups clearly defined by the content of cytosine which averaged 38·2% (molar % of total bases) in one group and 41·6% in the other. Within each group there are no significant differences between the animo acid compositions of the proteins but there are a number of marked differences between the groups. Likewise the amounts of the mono-, di- and most of the tri-nucleotides produced by ribonuclease digestion showed many differences between the groups whereas within each group smaller but statistically significant differences were found.

WCMV has a base ratio similar to, and TCV a base ratio very different from those of TYMV. Many large differences in amino acid composition and base sequence were found both between these two viruses and the six strains of TYMV.

No close correlations between base sequence and amino acid composition were found but differences in base composition and sequence were accompanied by marked differences in amino acid composition.

1. Introduction

Plant viruses offer an opportunity for studying the relationships between their RNA and protein components because each virus particle contains at least enough RNA to code for the sequence of amino acids in its protein coat. In addition, many of them can be isolated in pure form in gram quantities. The results of such an investigation should have some bearing on our understanding of the way in which genetic material can code for the synthesis of specific proteins.

Although it is possible to determine the sequence of amino acids in plant virus proteins, as has already been done for tobacco mosaic virus (Tsugita *et al.*, 1960; Anderer, Uhlig, Weber & Schramm, 1960), the determination of the nucleotide sequence in the associated RNA is much more difficult. Methods have been devised for the stepwise removal of nucleotides from the RNA chain (Whitfeld & Markham, 1953; Whitfeld, 1954; Brown, Fried & Todd, 1953; Yu & Zamecnik, 1960; Khym & Cohn, 1961) but so far they have only been used on short-chain oligonucleotides.

† Present address: Department of Agricultural Chemistry, Waite Agricultural Research Institute, University of Adelaide, South Australia.

‡ Abbreviations used: turnip yellow mosaic virus, TYMV; wild cucumber mosaic virus, WCMV; turnip crinkle virus, TCV.

Further, the 6000 mononucleotides present in a virus RNA molecule of molecular weight 2×10^6 would make any stepwise degradation a marathon task, if not an impossible one at the present time. However, some information can be gained on the type and quantity of base sequences which do occur by the use of enzymes which split the RNA chain specifically at certain phosphodiester bonds. Studies of this kind have been reported for the degradation of virus and yeast RNA by pancreatic ribonuclease by Markham & Smith (1952), Volkin & Cohn (1953), Reddi (1959*a*,*b*), Rushizky & Knight (1960) and Staehelin (1961).

This paper presents results on the amino acid composition, the base ratios of the RNA and the sequence of bases, as indicated by digestion of RNA with pancreatic ribonuclease, for a number of plant viruses. These include turnip yellow mosaic virus, wild cucumber mosaic virus and turnip crinkle virus. A preliminary account of some of this work has already been given (Symons, Rees & Markham, 1962).

2. Materials and Methods

Origin of virus strains

Six strains of TYMV have been used in the current investigation. Type strain is believed to be the one originally described by Markham & Smith (1949). Rademacher strain was isolated from a turnip, *Brassica rapa* var. *rapifera*, sent to this laboratory by Professor B. Rademacher, Stuttgart, Germany. Honesty strain was originally isolated from honesty plants (*Lunaria annua*) growing in a garden near Bristol (Broadbent, 1957). Cauliflower strain was isolated from an infected cauliflower (*B. oleracea* var. *botrytis*) and was probably obtained from a natural field infection in brassica crops in Northern England (Croxall, Gwynne & Broadbent, 1953). Rothamsted strain was obtained from Rothamsted Experimental Station, England, but details of its origin are not known. Denmark strain was obtained from H. Rønde-Kristensen, Denmark, who isolated it from a natural field infection of swede turnip, *B. napo-brassica*. WCMV was obtained from Dr. P. Kaesberg, Madison, Wisconsin, U.S.A. TCV was obtained from Dr. L. Broadbent (Broadbent & Heathcote, 1958).

Virus preparation

TYMV and TCV were grown in Chinese cabbage plants, *B. chinensis* var. *Pe Tsai*, and WCMV in pumpkin plants, *Cucurbito pepo*. Infected plants were ground in a commercial meat grinder and the virus isolated from the sap by differential centrifugation. In the case of TYMV and WCMV, the sap was first clarified with 20% ethanol and the virus precipitated with ammonium sulphate (Markham & Smith, 1949). All virus preparations were checked for purity in the Spinco model E ultracentrifuge before use.

Preparation of virus RNA and protein

TYMV. This virus was split into RNA and protein with 67% acetic acid by volume (Bawden & Pirie, 1937; Fraenkel-Conrat, 1957). The insoluble RNA was spun off and the pellet washed 3 times with 67% acetic acid and 3 times with acetone and allowed to dry at room temperature. The protein was precipitated from the supernatant fluid with 6 vol. acetone and the pellet washed with acetone and ether and dried at about 40°C in a stream of air.

WCMV. One lot of RNA and protein was prepared with 67% acetic acid as above, but there was incomplete splitting of the virus. Complete cleavage of the virus was obtained by heating solutions of virus in 2 M-KCl at 100°C for 10 min (Markham, unpublished experiments) under which conditions the protein precipitates and can be centrifuged off. The precipitated protein was washed with water, acetone and ether. The RNA was isolated from the supernatant by precipitation with 2 to 3 vol. ethanol and purified further by repeated precipitation with ethanol after being dissolved in water. Amino acid analyses (see below) were done on the protein isolated as above, on undegraded virus and on the top component of the virus which was isolated by density gradient centrifugation in sucrose (Brakke, 1951, 1953).

TCV. TCV is not degraded by acetic acid. Protein and RNA were prepared by heating the virus in 2 M-KCl as described for WCMV. For another preparation of TCV protein, the virus was degraded with 0·2% sodium dodecylsulphate at room temperature (Sreenivasaya & Pirie, 1938). After the addition of excess KCl, the insoluble potassium dodecylsulphate was spun off, the solution saturated with urea and passed through a charcoal column to remove the nucleic acid. The effluent, containing the protein, was dialysed salt-free in the cold and freeze-dried.

Base ratios of virus RNA

These were estimated by the method of Smith & Markham (1950) on RNA prepared as described above.

Ribonuclease digestion of RNA

Pancreatic ribonuclease (L. Light & Co., England, 4 times crystallized, salt and protease-free) was dissolved in water at a concentration of 5 mg/ml. and placed in a boiling-water bath for 3 min to remove any heat-labile contaminating enzymes. RNA was incubated, at a pH of about 7·5 with one-quarter its weight of ribonuclease, at 60°C for 2 hr. A small amount of bromothymol-blue was added as internal indicator and the pH kept at about 7·5 by the addition of N-ammonia. Buffered solutions were not used because Staehelin (1961) has shown that the presence of salt enhances the action of contaminating nucleases during the ribonuclease digestion of RNA from tobacco mosaic virus.

At the end of the digestion, the small amount of insoluble material was centrifuged off and the digest stored at −15°C. The optical density of each digest was determined at 260 mμ after suitable dilution with 0·05 M-ammonium bicarbonate buffer, pH 8·5. The total phosphate content of digests was determined by the method of Allen (1940) and was taken as a measure of the total bases present.

Column chromatography of ribonuclease digests

In general, the procedure of Staehelin (1961) using diethylaminoethyl cellulose (DEAE–cellulose) has been used. Before each run the DEAE–cellulose (Eastman Organic Chemicals, New York; capacity not stated) was thoroughly washed on a Büchner funnel with 0·5 N-KOH, water and finally 0·8 M-ammonium bicarbonate, pH 8·5, and then suspended in this buffer. It was packed into a column (1·3 cm diameter × 50 cm long) in 4 times its final volume of buffer at a pressure of 8 lb./in.^{2}. Ammonium bicarbonate, 0·5 M, pH 8·5, was pumped through the column for 2·5 to 4 hr at the rate of about 0·9 ml./min with a stainless steel piston pump (DCL Micro Pump, The Distillers Co. Ltd., Surrey, England). After equilibration, 1·0 ml. of ribonuclease digest of $O.D._{260m\mu}$ about 200 (1 cm cell) was added to the column and was eluted by an increasing concentration of ammonium bicarbonate obtained by means of a three-chambered apparatus (Peterson & Sober, 1959). Each chamber contained 470 ml. of ammonium bicarbonate, pH 8·5, at the following concentrations; chamber one, 0·05 M; chamber two, 0·25 M; chamber three, 1·2 M. The effluent from the column was run through a continuous-flow cell (0·5 cm light path) in a Cary 14 recording spectrophotometer and the optical density at 260mμ recorded at a chart speed of 0·14 cm/min. Fractions of about 4·5 ml. were collected at the rate of 0·9 ml./min, the time at which the fractions were changed being recorded on the chart with the wavelength-marking pen of the spectrophotometer which had been suitably connected to the fraction collector. After the run, the fractions of each peak were pooled, made to a known volume and the optical density read at 260 mμ in a 1 cm cell. Blank values, measured whenever the recording on the chart returned to the base line, increased slowly during the course of a run and usually varied between an optical density of 0·01 to 0·025.

Complete recovery of the material added to the column was obtained, the final few percent being eluted with 1·2 M-ammonium bicarbonate. The run was carried out overnight at 20 to 25°C and was complete in about 20 hr. Reasonable temperature control is essential for good fractionation. Further, since the pH of ammonium bicarbonate solutions varies markedly with temperature and increases slowly with time, it is necessary just prior to use to adjust the buffer solutions to the required pH at the temperature at which chromatography is to be carried out.

Satisfactory resolution of some compounds was not obtained during column chromatography and the unresolved peaks were separated by other means to determine the relative amounts of each compound present.

ApApUp† was separated from GpApCp and ApGpCp by paper electrophoresis at pH 3·5 (Markham & Smith, 1952) for the results with TYMV RNA. With WCMV and TCV RNA, all three compounds were separated in the ammonium sulphate solvent of Markham & Smith (1951) except that M-potassium phosphate pH 7 was substituted for M-sodium acetate. ApGpUp, GpGpCp and ApApApCp were separated either by electrophoresis at pH 3·5 or by chromatography with ammonium sulphate.

Results have been expressed as molar amounts of each compound found as a percentage of total nucleotides added to the column. Extinction coefficients as determined by Staehelin (1961) have been used.

Identification of products of ribonuclease digestion

Pooled samples of each peak eluted from the DEAE–cellulose column were taken to dryness on a rotary evaporator at about 40°C and the residue repeatedly dissolved in water and taken to dryness until all the ammonium bicarbonate had been removed. Identification of compounds was carried out by various means: ultraviolet spectra, paper electrophoresis (Markham & Smith, 1952), paper chromatography in various solvents, acid hydrolysis to give base ratios (Smith & Markham, 1950) and sequence determination of trinucleotides by the method of Whitfeld (1954) as modified by Yu & Zamecnik (1960). Cyclic adenylic and guanylic acids were prepared by the method of Smith, Moffat & Khorana (1958) for comparison with those found.

Amino acid composition of virus proteins

Samples of approximately 12 mg of each of the proteins of the virus strains prepared as described above, before or after oxidation with performic acid (Hirs, 1956), were hydrolysed for 24 hr by refluxing with constant boiling HCl. The excess HCl was removed by evaporation in a rotary evaporator at 40°C and the protein hydrolysate was redissolved in 15 ml. of pH 2·2 citrate buffer. Suitable samples were analysed for constituent amino acids as described by Spackman, Stein & Moore (1958). The tryptophan analyses were done by the method of Spies & Chambers (1949).

3. Results

Base ratios of virus RNA

In Table 1 are given the base ratios (mean values as moles% ± standard error for 4 to 6 estimations) for the RNA from six strains of TYMV and one strain each of WCMV and TCV. The most noticeable feature is the very high content of cytosine in TYMV and WCMV which varies from 37·9 to 42·1% of the total bases. It can be seen that the six strains of TYMV fall into two groups clearly defined by their content of cytosine, averaging 38·2% in group 1 and 41·6% in group 2. In addition the content of uracil in group 2 is lower than in group 1. Although WCMV has a similar content of cytosine and guanine to the TYMV strains of group 2, the contents of adenine and uracil are clearly different. The base ratios found for WCMV are similar to those published by Yamazaki & Kaesberg (1961*b*). The results for TCV have been included as an example of a virus which has chemical characteristics markedly different to those of TYMV and WCMV.

Statistical analysis of results

The results for the base ratios and for the products of ribonuclease digestion (see below) of the TYMV RNA have been analysed in two ways. The measured values for

† The nomenclature of Heppel, Whitfeld & Markham (1955) is used throughout this paper. G, A, C and U represent the nucleosides guanosine, adenosine, cytidine and uridine. The small p to the right of a nucleoside represents a 3′-phosphate to the left, a 5′-phosphate.

any one compound were compared in two strains by use of Student's t test for small samples using the standard errors of the means, either as calculated for each strain and presented in the Tables, or by using a mean value calculated from all the results for any one compound. The second method involves the combination of probabilities from tests of significance as described in section 21,1 of Fisher (1950). This method allows an assessment of the variability between two strains using all the measurements available and not just one as in the Student's t test. Thus, if a comparison of the results of, say, four compounds between two strains each show a difference which is significant at the 10% level as shown by the Student's t test, then taking the results together by using the second method of the combination of probabilities, it can be shown that these two strains show a difference which is significant at the 2% level.

TABLE 1

Base ratios of virus RNA

		Guanine	Adenine	Cytosine	Uracil
Strains of TYMV					
Group 1	type	17·2 ± 0·17	22·4 ± 0·11	38·3 ± 0·19	22·1 ± 0·17
	Rademacher	17·2 ± 0·14	22·8 ± 0·14	37·9 ± 0·24	22·1 ± 0·30
	honesty	17·5 ± 0·16	22·4 ± 0·17	38·6 ± 0·17	21·6 ± 0·28
Group 2	cauliflower	16·7 ± 0·11	21·4 ± 0·27	42·1 ± 0·34	19·8 ± 0·08
	Rothamsted	16·7 ± 0·14	21·3 ± 0·03	41·7 ± 0·23	20·3 ± 0·14
	Denmark	16·5 ± 0·16	21·6 ± 0·10	41·1 ± 0·11	20·7 ± 0·21
WCMV		16·4 ± 0·16	17·0 ± 0·04	41·0 ± 0·14	25·6 ± 0·16
TCV		27·8 ± 0·15	26·1 ± 0·12	23·7 ± 0·15	22·4 ± 0·19

Values given are in moles % and are the mean values ± standard error for 4 to 6 estimations.

As the number of estimations of any one compound in each strain of TYMV varied between three and eight, it is considered that the pooled estimate of the standard error of the mean for each compound is more realistic than the standard error as measured for each compound in each strain and it is this standard error which is used in all future discussion. However, similar results were obtained whichever standard error was used.

A comparison of the base ratios of the TYMV strains of group 1 as given in Table 1 by the Student's t test has shown no significant difference between the strains at the 5% level or less. However, a combination of probabilities showed that Rademacher differed from honesty at the 3% level but that there was no difference between the other strains. In group 2, the only significant differences were between the cauliflower and Denmark strains. The cytosine and uracil content of these two strains each differed at the 2% level or less. A combination of probabilities showed a difference between the two strains at the level of less than 1% (Table 4).

Ribonuclease digestion of virus RNA

The molecular weight of the RNA of TYMV, WCMV and TCV is about 2×10^6 (Markham, 1951; Haselkorn, 1962; Yamazaki & Kaesberg, 1961*b*; Haselkorn, Rees & Markham, unpublished results) and therefore each molecule of RNA contains about 6000 nucleotide residues. Some idea of the sequence of bases which occur in the RNA can be obtained by digesting the RNA with pancreatic ribonuclease which

splits the phosphodiester links of the RNA chain between the 3′-phosphate of a pyrimidine nucleotide residue and the 5′-position of the neighbouring nucleotide (Markham & Smith, 1952). Thus, oligonucleotides of varying chain length are liberated, each consisting of 0, 1 or more purine nucleotide residues terminated by a pyrimidine nucleoside 3′-phosphate.

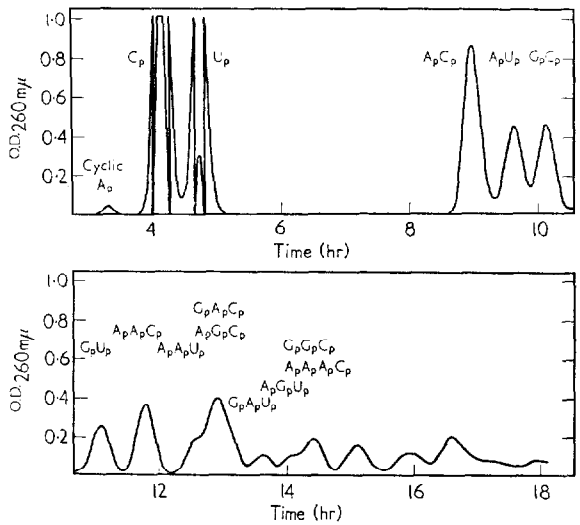

FIG. 1. Chromatography of pancreatic ribonuclease digest of TYMV RNA. Tracing of optical density of effluent from a DEAE–cellulose column as obtained with a Cary recording spectrophotometer. For the Cp and Up peaks, the optical density increased above 1·0 and the optical density tracing between 1·0 and 2·0 is given by the innermost lines of these two peaks. For further details see text.

The digestion of RNA with ribonuclease and the fractionation of the digestion products on a DEAE–cellulose column have already been described in the experimental section. Figure 1 shows a tracing of a chromatogram obtained on a Cary 14 recording spectrophotometer during one of the runs using RNA from TYMV. It can be seen that the two possible mononucleotides and the four possible dinucleotides are readily separated. Although only two of the eight possible trinucleotides have separated free from other compounds, it is possible to estimate five more trinucleotides and one of the sixteen possible tetranucleotides. The peak containing the trinucleotides ApApCp, GpApCp and ApGpCp and that containing ApGpUp, GpGpCp and ApApApCp can be separated into their components by chromatography in buffered ammonium sulphate. In the case of the results with TYMV, ApApUp was separated by electrophoresis from the two isomers GpApCp and ApGpCp which move together so that results for these two isomers are given together. The only other trinucleotide, GpGpUp, was eluted with the higher nucleotides and was therefore not estimated.

The results for TYMV presented in Table 2 again show that the six strains clearly fall into the same two groups mainly on the basis of differences in the amounts of Cp, Up, ApUp, ApApCp, ApApUp and GpApUp. The main feature of the results for TYMV and WCMV is the high proportion of mononucleotides which are liberated, i.e. the proportion of total bases existing as two or more pyrimidines side by side varies from 39·9 to 42·6% for TYMV strains and 49·3% for WCMV. By contrast, the

TABLE 2

Oligonucleotides in pancreatic ribonuclease digests of virus RNA

Nucleotide	Strains of TYMV						WCMV	TCV
	Type	Rademacher	Honesty	Cauliflower	Rothamsted	Denmark		
Cp	26·5 ±0·17	25·5 ±0·13	26·2 ±0·18	29·0 ±0·08	29·2 ±0·08	28·0 ±0·14	31·0 ±0·17	13·5 ±0·08
Up	14·6 ±0·18	14·4 ±0·10	14·6 ±0·09	13·3 ±0·06	13·4 ±0·18	13·9 ±0·10	18·3 ±0·17	12·6 ±0·14
ApCp	10·1 ±0·08	10·3 ±0·10	10·0 ±0·08	9·95±0·05	10·0 ±0·07	10·6 ±0·08	6·78±0·08	6·46±0·06
ApUp	5·60±0·08	5·28±0·14	5·72±0·14	4·69±0·05	4·85±0·07	4·70±0·09	4·48±0·07	5·12±0·07
GpCp	6·30±0·09	5·88±0·08	6·64±0·16	6·35±0·11	6·60±0·06	6·41±0·13	7·13±0·10	5·35±0·17
GpUp	4·12±0·10	3·80±0·07	4·30±0·13	3·39±0·15	3·68±0·04	3·96±0·09	3·98±0·06	5·35±0·09
ApApCp	3·97±0·06	3·60±0·24	3·78±0·22	4·37±0·01	4·50±0·02	4·36±0·03	2·12±0·03	3·46±0·06
ApApUp	1·45±0·04	1·53±0·12	1·47±0·10	2·12±0·17	2·08±0·08	1·76±0·07	1·32±0·04	2·52±0·02
GpApGp	6·72±0·08	6·15±0·13	6·51±0·07	6·57±0·07	6·35±0·09	6·52±0·20	2·66±0·04	2·60±0·06
ApGpCp							2·31±0·04	3·50±0·05
GpApUp	1·77±0·04	1·71±0·03	1·75±0·02	1·22±0·01	1·38±0·03	1·15±0·04	1·64±0·05	2·60±0·08
ApGpUp	1·31±0·04	1·56±0·08	1·60±0·11	1·11±0·10	1·13±0·03	1·39±0·09	1·21±0·06	2·72±0·03
GpGpCp							1·60±0·08	2·15±0·05
ApApApCp							1·01±0·04	0·74±0·14
Sum	82·4	79·7	82·6	82·1	83·2	82·8	85·5	85·5

Values given are mean values of mononucleotide residues as a % of total nucleotides added to the chromatography column together with the standard errors of the mean. Any one figure has been calculated from the results of from 3 to 8 separate estimations.

value for TCV is only 26·1%. These figures represent from 66·5 to 68·7% of the total pyrimidines for TYMV, 74·1% for WCMV and 56·2% for TCV. These values are roughly proportional to the total pyrimidine content of the RNA of each virus, as would be expected if the bases were distributed in a more or less random fashion along the RNA chain. Similar calculations can be made for the values given for the other digestion products so that considerable numerical information on the distribution of certain sequences along the RNA chain can be gained (see e.g. Reddi, 1959*b*).

The sum totals of the digestion products listed in Table 2 show a remarkable similarity for the eight viruses studied. In seven of these, the sum of the mono- and dinucleotides and six of the possible eight trinucleotides varies between 82·1 and 83·2% of the total and is 79·7% for the Rademacher strain of TYMV.

TABLE 3

Ratio of found to random distribution of oligonucleotides in ribonuclease digests of virus RNA

Nucleotide	Molar ratios found/random in ribonuclease digests							
	Strains of TYMV						WCMV	TCV
	Type	Rademacher	Honesty	Cauli-flower	Rotham-sted	Denmark		
Cp	1·15	1·12	1·13	1·11	1·13	1·10	1·14	1·24
Up	1·10	1·08	1·12	1·08	1·06	1·09	1·07	1·22
ApCp	0·97	1·00	0·96	0·89	0·91	0·97	0·73	1·13
ApUp	0·94	0·87	0·98	0·89	0·90	0·85	0·77	0·95
GpCp	0·79	0·75	0·82	0·73	0·76	0·76	0·80	0·88
GpUp	0·90	0·83	0·95	0·83	0·88	0·94	0·71	0·93
ApApCp	1·14	1·01	1·08	1·22	1·28	1·23	0·90	1·55
ApApUp	0·72	0·74	0·75	1·26	1·22	0·98	0·89	1·19
GpApCp / ApGpCp	1·26	1·15	1·19	1·18	1·15	1·20	1·16 / 1·01	1·14 / 1·47
GpApUp	1·15	1·10	1·14	0·93	1·03	0·84	1·15	1·16
ApGpUp	0·85	1·00	1·05	0·85	0·84	1·01	0·85	1·21
GpGpCp	—	—	—	—	—	—	0·73	0·85
ApApApCp	—	—	—	—	—	—	1·88	0·95

For the results for the found distribution of nucleotides, the mean values in Table 2 have been used. The distributions expected on a random basis have been calculated from the base ratios given in Table 1.

It will be noted that some of the positional isomers are not present in equal amounts, which indicates that there is a non-random distribution of bases along the RNA chain. For example, in some strains of TYMV and in WCMV, GpApUp differs in amount to ApGpUp whereas, in TCV and in WCMV, GpApCp is produced in different amounts to ApGpCp. On the basis of the base ratios of Table 1, the theoretical amounts of the oligonucleotides expected from a ribonuclease digest have been calculated assuming a random distribution of bases along the RNA chain (Staehelin, 1961). The values given in Table 2 have been expressed as a fraction of these random values and the results of these calculations are given in Table 3. It can be seen that the experimental values show marked variations from random. For TYMV there is a variation of from 0 to 28% on either side of the random values, for WCMV variations of from −29% to +88%, and for TCV variations from −15% to +55%.

The results of Table 3 also show some variations from random which are consistent in all the strains examined. For example, Cp and Up are produced in amounts in excess of those predicted from a random distribution of bases. For the strains of TYMV and for WCMV the molar ratios of "found" to "random" for the mononucleotides are closely similar whilst the values are higher for TCV. All values for the trinucleotides GpApCp and ApGpCp also show a positive variation from random. Further, except for one case each in the TYMV Rademacher strain and in TCV, all four dinucleotides show a consistent negative variation from random. However, the meaning of any consistent variation from random is not known.

TABLE 4

Statistical comparison of results obtained with RNA from six strains of turnip yellow mosaic virus

Strains compared	Probability p	
	Base ratios	Ribonuclease digests
Type *v.* Rademacher	0·71	0·0001
Type *v.* honesty	0·36	0·18
Rademacher *v.* honesty	0·03	0·001
Cauliflower *v.* Rothamsted	0·32	0·62
Cauliflower *v.* Denmark	0·007	0·001
Rothamsted *v.* Denmark	0·21	0·001

Values are probabilities p that the two strains compared are the same, calculated by combination of probabilities from t tests of significance.

The results on the ribonuclease digestion products clearly show, as did the base ratios, that WCMV and TCV are different from all strains of TYMV and from each other. Also, it is clear that there are marked differences between the two groups of TYMV strains, especially when we remember that a difference of 1% represents about sixty mononucleotide residues. However, within each group the results are similar and require a statistical examination to show where definite differences occur. It must be emphasized that strains that do not differ statistically could well vary appreciably in their nucleotide sequence where the variations are not sufficiently large to be reflected in the experimental results.

Application of the Student's t test for small samples to the eleven compounds estimated for each of the six strains of TYMV showed from 0 to 6 differences at the 5% level or less for the three pairs of strains which could be compared within each group. The results are: type *v.* Rademacher—6 differences, type *v.* honesty—2, Rademacher *v.* honesty—4, cauliflower *v.* Rothamsted—0, cauliflower *v.* Denmark—4, Rothamsted *v.* Denmark—3. The results obtained by a combination of probabilities from these tests of significance (see above) are given in Table 4. It can be seen that the differences between four of the pairs are highly significant and it is considered, therefore, that there are marked differences in the sequence of bases in the RNA of the strains of each pair. The results suggest that the cauliflower and Rothamsted strains are similar but minor differences could still occur, the detection of which would require finer analytical techniques. The position with the type and honesty strains is less clear. The Student's t test showed two differences, one significant at the 5%

and the other at the 3% level whereas the combination of probabilities only showed a difference at the 18% level. It is therefore considered possible that these two strains are different.

Presence of cyclic adenylic and guanylic acids in ribonuclease digests

In view of the specificity of pancreatic ribonuclease, it was surprising to find small but consistent amounts of 2',3'-cyclic adenylic and guanylic acids in the ribonuclease digestion of RNA from all strains of virus tested. As can be seen from Fig. 1, cyclic adenylic acid is clearly separated from cytidylic acid on column chromatography and can be readily estimated. However, cyclic guanylic acid is eluted with uridylic acid and its amount was therefore determined by subsequent chromatography in the 70% isopropanol–ammonia solvent of Markham & Smith (1952). A summary of the amounts of these cyclic compounds found is shown in Table 5. In all cases, cyclic adenylic acid was formed in similar amounts and was in excess of cyclic guanylic acid. There are fewer results for cyclic guanylic acid but larger quantities were produced with WCMV than with TCV.

TABLE 5

Amounts of 2',3' cyclic adenylic and guanylic acids in ribonuclease digests of virus RNA

Virus	Cyclic adenylic acid	Cyclic guanylic acid
TYMV	0·32 (0·28–0·45) (28)	—
WCMV	0·40 (0·30–0·47) (7)	0·23 (0·20–0·26) (3)
TCV	0·31 (0·25–0·37) (5)	0·052 (0·006–0·11) (4)

Values as moles % of total nucleotides. Mean values together with first figures in parenthesis, range of values obtained; second figure in parenthesis, number of estimations.

Staehelin (1961) has reported the production of considerable quantities of cyclic adenylic and guanylic acids during the digestion of virus RNA with pancreatic ribonuclease in the presence of salt. In the absence of salt, however, he says that these cyclic compounds were not formed but an examination of his published chromatograms would suggest that small amounts were still present. Staehelin explained this effect of salt as a stimulation of contaminating plant nucleases. Although this is also a possibility in our own case, the consistency of production of cyclic adenylic acid in a large number of digests of RNA obtained from many virus preparations made throughout the year would suggest that some other explanation is required.

Amino acid composition of virus proteins

The amino acid compositions of the viruses under study are given in Table 6 and are presented as the nearest whole number of residues, assuming a molecular weight of about 20,000 for the protein subunit, calculated from two or more analyses both before and after performic acid oxidation.

It can be seen from the results of Table 6 that the proteins of the six strains of TYMV clearly fall into the same two groups as found also for the results obtained on their RNA. There are a number of marked differences between the two groups, especially in the content of lysine, histidine, arginine, aspartic acid, threonine, serine

TABLE 6

Amino acid composition of virus proteins

Amino acid	Strains of TYMV						WCMV			TCV
	Group 1			Group 2						
	Type	Honesty	Rademacher	Cauliflower	Rothamsted	Denmark	Bottom component	Top component	Whole virus	
Lysine	7	7	7	4	4	4	9	9	8	12
Histidine	3	3	3	5	5	5	3	3	3	1
Arginine	3	3	3	6	6	6	5	5	5	9
Tryptophan	2	2	2	2	2	2	1	1	1	4
Aspartic acid	11	11	11	17	17	17	15	15	15	14
Threonine†	26	26	26	20	19	20	13	13	13	14
Serine†	16	16	16	20	19	20	26	26	25	12
Glutamic acid	15	14	14	15	14	15	11	11	11	16
Proline	20	20	20	20	21	20	18	19	19	9
Glycine	8	8	8	7	7	7	9	9	13	15
Alanine	15	14	14	13	13	13	16	16	16	17
Half cystine	4	4	4	5	5	5	2	2	—	1
Valine	14	15	13	14	13	14	12	12	13	12
Methionine	4	4	4	4	4	4	1	1	1	2
Isoleucine	15	15	17	13	13	14	13	13	13	5
Leucine	18	17	17	19	19	18	24	24	24	11
Tyrosine	3	3	3	3	3	3	3	3	3	4
Phenylalanine	5	5	5	3	3	3	8	8	8	6
Total amino acids	189	187	187	190	187	190	189	190	—	164
Molecular weight	20,000	19,900	19,900	20,400	20,100	20,400	20,100	20,200	—	17,800

Values are residues of amino acids per mole of protein. Results for TCV are from Haselkorn, Rees & Markham (unpublished results).

† Threonine and serine values have been corrected for losses during acid hydrolysis; 5% for threonine and 10·5% for serine.

and phenylalanine. Within each group the results are closely similar but the over-all amino acid analyses could conceal differences that might be made apparent by structural studies.

Analyses done on the whole virus, top component and protein isolated from the bottom component of WCMV have shown the same amino acid composition and are similar to the results obtained by Yamazaki & Kaesberg (1961*a*,*b*). The higher content of glycine obtained from the whole virus is probably due to the breakdown of the RNA purines during acid hydrolysis (Krüger, 1892; Wulff, 1893). Further, it can be seen that both WCMV and TCV have amino acid compositions which are appreciably different to those obtained for TYMV and for each other.

4. Discussion

It can be seen from the results presented that any marked differences in the base composition and sequence of virus RNA are accompanied by marked differences in the amino acid composition, and hence also sequence, of the virus protein. For the results obtained with the six strains of TYMV, it has not been possible to find any correlation between changes in the base sequence and the amino acid composition which would provide information on the way virus RNA codes for the synthesis of its protein coat. This is, perhaps, to be expected as, on the basis of the presently accepted concept of a code involving nucleotide triplets, only $189 \times 3 = 567$ bases would be required to code for a mole of virus protein; this represents only about 10% of the total bases present in an RNA molecule of molecular weight 2×10^6. One would, therefore, expect to find a correlation between virus protein and RNA only in those viruses where the molar ratio of bases to amino acids approaches three. Unfortunately, there are very few viruses known at present which come near to these requirements but the most promising is the satellite component of tobacco necrosis virus (Kassanis & Nixon, 1961) which has a ratio of bases to amino acids of five (Reichmann, Rees & Markham, 1962; Reichmann, Rees, Symons & Markham, 1962).

The results for the six strains of TYMV are of further interest as an example of the chemical differentiation of virus strains. A difference in the amino acid composition or sequence of the virus protein clearly indicates at least a corresponding difference in the base sequence of the RNA. However, it is possible to have two strains with the same protein composition but different biological properties (and perhaps different base sequence in most of the virus RNA) as has already been found for a number of chemically evoked mutants of tobacco mosaic virus (Tsugita & Fraenkel-Conrat, 1962; Wittmann, 1961).

Since it is conceivable that two virus strains which show no obvious biological differences can still vary considerably in the base sequence in their RNA, the ultimate comparison of two strains must lie in the base sequence of their RNA. The results obtained for the six strains of TYMV show that the analysis of pancreatic ribonuclease digests of virus RNA can show up reasonably small differences between strains. However, since a difference of 1% represents sixty nucleotide residues in an RNA molecule of molecular weight 2×10^6, a difference which is just significant in the present work, still represents quite a few base changes. An additional help in yielding information on base sequence would be the enzyme ribonuclease T_1 which has a similar action to pancreatic ribonuclease but only releases nucleotide fragments terminated by a 3′-guanylic acid (Sato & Egami, 1957). Preliminary work on the use of this enzyme has been reported by Miura & Egami (1960), Reddi (1960) and

Rushizky, Sober & Knight (1962). The real advantage of this enzyme is that the frequency of, say, the isolated dinucleotide ApGp at once gives the frequency of the trinucleotide GpApGp. In the case of pancreatic ribonuclease, however, an isolated dinucleotide, say ApUp, has had its origin in both UpApUp and CpApUp.

The ultimate usefulness of these enzymes will be determined by the development of techniques for the separation of the tetra- and higher oligonucleotides. Pancreatic ribonuclease, for example, produces sixteen tetranucleotides and the T_1 enzyme twenty-seven, but it is not possible at present to obtain more than a rough fractionation of these. Unfortunately it is in these and the higher nucleotides that really small differences in base sequence are likely to be found.

Although a complete description of a virus strain must include details of its chemical structure, it is of interest to consider the type of differences which are likely to occur both within and between virus strains. It might be possible for a single base difference in virus RNA to result in a difference in the amino acid composition of the virus protein and possibly its serological behaviour. Such a difference in the RNA might well have no effect on the phenotypic expression of the virus nor could it be detected analytically. On the other hand, it is conceivable that changes could occur in the virus RNA which are unaccompanied by changes in the protein and in phenotypic expression but which could be detected by analysis of enzyme digests of the RNA. One should therefore be prepared to find that any normally accepted strain of virus, such as the type strain of TYMV, consists of a whole population of particles which have the protein coat, and that part of the RNA chain which codes for it the same or very similar, but having many variations in the remaining part of the RNA chain. Such a situation would have a distinct advantage for the survival of the virus under varying environmental conditions and on transfer from one host to another because the most effective population of virus particles would soon be selected out.

An important problem in plant viruses is the determination of the function of that part of the RNA which does not code for virus protein. By analogy with recent work on bacteriophage (see e.g. Flaks & Cohen, 1957; Kornberg, Zimmerman, Kornberg & Josse, 1959; Bello, van Bibber & Bessman, 1961; Koerner, Smith & Buchanan, 1960) it would seem that some or all of this extra RNA codes for the synthesis of enzymes and other materials which are necessary for the multiplication of the virus in the plant host. The only evidence bearing on this is the observation of Tsugita, Fraenkel-Conrat, Nirenberg & Matthaei (1962) that TMV RNA, acting as messenger RNA in a cell-free system incorporating amino acids, can stimulate the incorporation of amino acids which are not present in the virus protein. Circumstantial evidence that this extra RNA also has a function is the finding of Johnson & Markham (1962) that the polyamine, $NH_2(CH_2)_3NH(CH_2)_3NH_2$, present in all plant viruses so far examined, is not present in uninfected host plants.

It has already been noted above that TYMV RNA contains about 10 times as much RNA as is necessary to code for its own protein, on the assumption that three bases code for each amino acid. On the basis of the triplet codes at present published (Matthaei, Jones, Martin & Nirenberg, 1962; Speyer, Lengyel, Basilio & Ochoa, 1962) and on the data presented in this paper, it is possible to make interesting calculations on the distribution of bases along the virus RNA chain. However, although such calculations are important, it is considered that the presentation of these calculations is unwarranted because of the changes in our knowledge of the genetic code which are likely to occur in the near future (see e.g. Reichmann, Rees, Symons & Markham, 1962).

This work was partly supported by the U.S. Public Health Service. R. H. S. was in receipt of a C.S.I.R.O. Overseas Studentship. The authors thank J. G. Rowell, A.R.C. Statistics Group, School of Agriculture, University of Cambridge, for statistical advice on the interpretation of results.

REFERENCES

Allen, R. J. L. (1940). *Biochem. J.* **34**, 858.
Anderer, F. A., Uhlig, H., Weber, E. & Schramm, G. (1960). *Nature*, **186**, 922.
Bawden, F. & Pirie, N. W. (1937). *Proc. Roy. Soc.* B, **123**, 274.
Bello, L. J., van Bibber, M. J. & Bessman, M. J. (1961). *J. Biol. Chem.* **236**, 1467.
Brakke, M. K. (1951). *J. Amer. Chem. Soc.* **73**, 1847.
Brakke, M. K. (1953). *Arch. Biochem. Biophys.* **45**, 275.
Broadbent, L. (1957). In *Investigation of Virus Diseases of Brassica Crops*, p. 76. Agricultural Research Council, Cambridge University Press.
Broadbent, L. & Heathcote, G. D. (1958). *Ann. App. Biol.* **46**, 585.
Brown, D. M., Fried, M. & Todd, A. R. (1953). *Chem. & Ind.* p. 352.
Croxall, H. E., Gwynne, D. C. & Broadbent, L. (1953). *Plant Path.* **2**, 122.
Fisher, R. A. (1950). In *Statistical Methods for Research Workers*, 11th edition, section 21.1, p. 99. London: Oliver & Boyd.
Flaks, J. G. & Cohen, S. S. (1957). *Biochim. biophys. Acta*, **25**, 667.
Fraenkel-Conrat, H. (1957). *Virology*, **4**, 1.
Haselkorn, R. (1962). *J. Mol. Biol.* **4**, 357.
Heppel, L. A., Whitfeld, P. R. & Markham, R. (1955). *Biochem. J.* **60**, 8.
Hirs, C. H. W. (1956). *J. Biol. Chem.* **211**, 611.
Johnson, M. W. & Markham, R. (1962). *Virology*, **17**, 276.
Kassanis, B. & Nixon, H. L. (1961). *J. Gen. Microbiol.* **25**, 459.
Khym, J. X. & Cohn, W. E. (1961). *J. Biol. Chem.* **236**, PC 9.
Koerner, J. F., Smith, M. S. & Buchanan, J. M. (1960). *J. Biol. Chem.* **235**, 2691.
Kornberg, A., Zimmerman, S. B., Kornberg, S. R. & Josse, J. (1959). *Proc. Nat. Acad. Sci., Wash.* **45**, 772.
Krüger, M. (1892). *Z. physiol. Chem.* **16**, 160.
Markham, R. (1951). *Disc. Faraday Soc.* **11**, 221.
Markham, R. & Smith, J. D. (1951). *Biochem. J.* **49**, 401.
Markham, R. & Smith, J. D. (1952). *Biochem. J.* **52**, 552, 558, 565.
Markham, R. & Smith, K. M. (1949). *Parasitology*, **39**, 330.
Matthaei, J. H., Jones, O. W., Martin, R. G. & Nirenberg, M. W. (1962). *Proc. Nat. Acad. Sci., Wash.* **48**, 666.
Miura, K. & Egami, F. (1960). *Biochim. biophys. Acta*, **44**, 378.
Peterson, E. A. & Sober, H. A. (1959). *Anal. Chem.* **31**, 857.
Reddi, K. K. (1959*a*). *Biochim. biophys. Acta*, **32**, 386.
Reddi, K. K. (1959*b*). *Proc. Nat. Acad. Sci., Wash.* **45**, 293.
Reddi, K. K. (1960). *Nature*, **188**, 60.
Reichmann, M. E., Rees, M. W. & Markham, R. (1962). *Biochem. J.* **84**, 86 P.
Reichmann, M. E., Rees, M. W., Symons, R. H. & Markham, R. (1962). *Nature*, **195**, 999.
Rushizky, G. W. & Knight, C. A. (1960). *Proc. Nat. Acad. Sci., Wash.* **46**, 945.
Rushizky, G. W., Sober, H. A. & Knight, C. A. (1962). *Biochim. biophys. Acta*, **61**, 56.
Sato, K. & Egami, F. (1957). *J. Biochem., Tokyo*, **49**, 1.
Smith, J. D. & Markham, R. (1950). *Biochem. J.* **46**, 509.
Smith, M., Moffat, J. G. & Khorana, H. G. (1958). *J. Amer. Chem. Soc.* **80**, 6204.
Spackman, D. H., Stein, W. H. & Moore, S. (1958). *Anal. Chem.* **30**, 1190.
Speyer, J. F., Lengyel, P., Basilio, C. & Ochoa, S. (1962). *Proc. Nat. Acad. Sci., Wash.* **48**, 441.
Spies, J. R. & Chambers, D. C. (1949). *Anal. Chem.* **21**, 1249.
Sreenivasaya, M. & Pirie, N. W. (1938). *Biochem. J.* **32**, 1707.
Staehelin, M. (1961). *Biochim. biophys. Acta*, **49**, 11, 20, 27.
Symons, R. H., Rees, M. W. & Markham, R. (1962). *Biochem. J.* **84**, 37 P.
Tsugita, A. & Fraenkel-Conrat, H. (1962). *J. Mol. Biol.* **4**, 73.

Tsugita, A., Fraenkel-Conrat, H., Nirenberg, M. W. & Matthaei, J. H. (1962). *Proc. Nat. Acad. Sci., Wash.* **48**, 846.
Tsugita, A., Gish, D. T., Young, J., Fraenkel-Conrat, H., Knight, C. A. & Stanley, W. M. (1960). *Proc. Nat. Acad. Sci., Wash.* **46**, 1463.
Volkin, E. & Cohn, W. E. (1953). *J. Biol. Chem.* **205**, 767.
Whitfeld, P. (1954). *Biochem. J.* **58**, 390.
Whitfeld, P. & Markham, R. (1953). *Nature*, **171**, 1151.
Wittmann, H. G. (1961). *Naturwiss.* **48**, 55.
Wulff, C. (1893). *Z. physiol. Chem.* **17**, 468.
Yamazaki, H. & Kaesberg, P. (1961*a*). *Biochim. biophys. Acta*, **53**, 173.
Yamazaki, H. & Kaesberg, P. (1961*b*). *Biochim. biophys. Acta*, **51**, 9.
Yu, C. & Zamecnik, P. C. (1960). *Biochim. biophys. Acta*, **45**, 148.